Problem 1

Calculate the Lagrangian $\mathcal{L}(x, \alpha)$.

$$\mathcal{L}(x, \alpha) = -(x_1 + x_2) + \alpha_1(x_1^2 + x_2^2 - 1)$$

Obtain the Lagrangian dual function $g(\alpha)$.

$$\begin{aligned}
\frac{\nabla_{\mathcal{L}(x,\alpha)}}{\partial x_1} &= 0 \Leftrightarrow \\
2\alpha_1 x_1 - 1 &= 0 \Leftrightarrow \\
x_1 &= \frac{1}{2\alpha_1}
\end{aligned}$$

$$\begin{aligned}
\frac{\nabla_{\mathcal{L}(x,\alpha)}}{\partial x_2} &= 0 \Leftrightarrow \\
2\alpha_1 x_2 - 1 &= 0 \Leftrightarrow \\
x_2 &= \frac{1}{2\alpha_1}
\end{aligned}$$

Solve the dual problem (Plug x^* in $g(\alpha)$).

$$\begin{aligned}
\frac{\nabla_{g(\alpha)}}{\partial \alpha} &= 0 \Leftrightarrow \\
\frac{1}{\alpha_1^2} - \frac{1}{2\alpha_1^2} - 1 &= 0 \Leftrightarrow \\
\alpha_1^2 &= \frac{1}{2} && \text{(by constraint } \alpha_i \geq 0\text{)} \\
\alpha_1 &= \frac{1}{\sqrt{2}}
\end{aligned}$$

Problem 2

- Similarities
  - Both algorithms try to solve the problem of binary classification by finding a decision boundary $w^\intercal x + b = 0$ that separates all datapoints x_i with label 1 from all datapoints x_j with label -1
- Differences
  - SVM has a closed form solution
  - SVM gives an unique solution (constrained optimization) by choosing a decision boundary s.t. it has a maximum margin to its nearest datapoints
  - Perceptron must be solved iteratively
  - Perceptron may have infinitely many correct solutions (if available) (unconstrained optimization)

Problem 3

By the formulation of the SVM problem we have

$$
\begin{aligned}
\text{minimize} \quad & f_0(w,b) = \frac{1}{2} w^\intercal w \\
\text{subject to} \quad & f_i(w,b) = y_i(w^\intercal x_i + b) - 1 \geq 0, \quad \text{for i = 1, ...,N}
\end{aligned} \tag{1}
$$

Clearly we can rewrite (1) to

$$f_i(w,b) = -y_i(w^\intercal x_i + b) + 1 \leq 0, \quad \text{for i = 1, ...,N}$$

By Slater's constraint qualification we have that the duality gap of the SVM problem is zero if $f_0(x)$, $f_1(x)$, ... $f_N(x)$ are convex and the constraints $f_1(x)$, ... $f_N(x)$ are affine. Clearly both assumptions are met, because $f_0(x)$ is simply the L_2-norm, which is convex and the constraints are linear functions in w shifted by an offset b, which makes them affine. Thus the duality gap is zero.

Problem 4

a). Let $\mathbf{X} \in \mathbb{R}^{n \times d}$, where each row is a datapoint $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{Y} \in \mathbb{R}^{n \times n}$, where each column is the label y_i of the i-th datapoint replicated n times and $\boldsymbol{\alpha} \in \mathbb{R}^n$ with α_i at position i.
By the Hadamard product we have

$$\mathbf{Q} = XX^\intercal \odot (Y \odot Y^\intercal)$$

$$g(\boldsymbol{\alpha}) = \boldsymbol{\alpha}^\intercal \mathbf{Q} \boldsymbol{\alpha}$$

b). TODO
c). TODO